Web aplikacija za aproksimaciju podataka metodom najmanjih kvadrata

Biondić, Marija

Undergraduate thesis / Završni rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:200:247706

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-04-20**

Repository / Repozitorij:

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I INFORMACIJSKIH TEHNOLOGIJA

Stručni studij

Web-aplikacija za aproksimaciju podataka metodom najmanjih kvadrata

Završni rad

Marija Biondić

Osijek, 2020.

SADRŽAJ

1.UVOD 2

2.TEORIJSKA PODLOGA 3

2.1.Jednostavna linearna regresija 3

2.1.1. Koeficijent determinacije (odlučnosti) R^2 6

2.2. Višestruka linearna regresija 8

2.2.1. Matrični pristup višestrukoj linearnoj regresiji 10

2.2.2. Polinomni regresijski modeli 11

2.3. Područja primjene 11

3. ALATI ZA IZRADU WEB-APLIKACIJE 12

3.1. VisualStudio Code 12

3.2. HTML I CSS 12

3.3. JavaScript 14

3.3.1. D3.js 15

3.4. FileAPI 17

3.5. XAMPP 18

4. PROGRAMSKA IMPLEMENTACIJA 19

5. ZAKLJUČAK 23

LITERATURA 24

SAŽETAK 25

ABSTRACT 26

PRILOZI 27

1.UVOD

Princip najmanjih kvadrata postavio je 1795. godine Carl Friedrich Gauss prilikom izučavanja kretanja nebeskih tijela, dok je prvi objavio algebarsku proceduru metode najmanjih kvadrata francuski matematičar Legendre 1805. godine.

Metoda najmanjih kvadrata je statistička metoda određivanja krivulje ili funkcije koja najbolje odgovara (engl. *line of best fit*) zadanom modelu. Metoda je definirana jednadžbom sa određenim parametrima promatranih podataka. Ova metoda je standardno korištena u regresijskoj analizi, koja se koristi za donošenje zaključaka o nizu slučajnih varijabli $Y_1, \ldots, Y_n$ koje ovise o nezavisnoj varijabli x.
Pri korištenju metode najmanjih kvadrata teško će se ili nikada dobiti krivulja koja savršeno opisuje zadane točke.

Cilj rada je opisati metodu najmanjih kvadrata te izrada web-aplikacije za rješavanje određenih zadataka.

Unutar drugog poglavlja obrađena je i prezentirana teorijska podloga statističkih i matematičkih modela na kojima se zasniva rad web-aplikacije.

U trećem poglavlju je napravljen pregled alata i tehnologija korištenih pri izradi aplikacije.

U četvrtome poglavlju prezentirane su funkcionalnosti te izgled i način rada aplikacije.

2.TEORIJSKA PODLOGA

Zbirka statističkih alata koji se koriste za modeliranje i istraživanje odnosa između varijabli, koje su povezane na nedeterministički način, naziva se regresijska analiza. Problemi ove vrste javljaju se često u mnogim granama tehnike i znanosti, stoga je regresijska analiza jedan od najčešće korištenih statističkih alata. Za ilustraciju analize se najčešće se koristi dijagram rasipanja (engl. *scatter diagram*).

Pregled dijagrama rasipanja pokazuje da, iako jednostavna krivulja neće proći točno kroz sve točke, postoji snažna naznaka da točke leže nasumično raspršene oko pravca. Stoga je razumna pretpostavka da je aritmetička sredina zavisne varijable *Y* povezana s *x* sljedećim pravocrtnim odnosom:

$$E(Y|x) = \beta_0 + \beta_1 x \tag{2-1}$$

gdje se nagib (β_1) i odsječak na osi (β_0) nazivaju regresijskim koeficijentima.

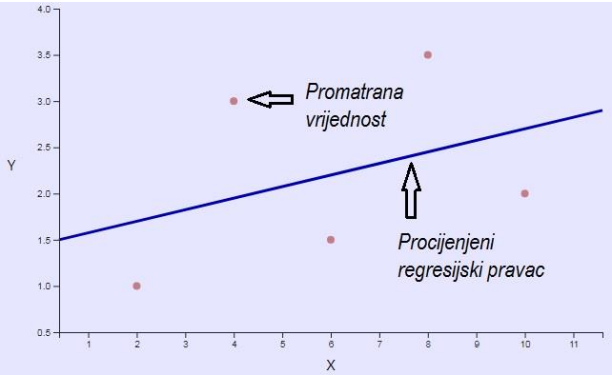
2.1.Jednostavna linearna regresija

Iako aritmetička sredina od *Y* je linearna funkcija od *x*, stvarna promatrana vrijednost *y* ne pada točno na pravac. Prikladan način da se ovo generalizira na probabilistički (vjerojatnostni) linearni model jest pretpostaviti da je očekivana vrijednost *Y* linearna funkcija od *x*, ali za fiksnu vrijednost *x*. Stvarna vrijednost *Y* se određuje funkcijom srednje vrijednosti (linearni model) plus slučajna pogreška, stoga vrijedi sljedeći izraz:

$$Y = \beta_0 + \beta_1 x + \in \tag{2-2}$$

gdje je $\in$ slučajna pogreška nepoznate varijance σ^2 . Takav model se naziva jednostavni linearni regresijski model jer ima samo jednu nezavisnu varijablu *x* i zavisnu varijablu *Y*. [1]

Pretpostavimo da imamo *n* parova promatranja (podataka/mjerenja) $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, slika 2.1 je tipičan dijagram rasipanja promatranih podataka i kandidat za procijenjeni (engl. *estimated*) regresijski pravac.



Sl. 2.1. *Odstupanje podataka od procijenjenog regresijskog pravca.*

Procijene od β_0 i β_1 trebale bi rezultirati pravcem koji se, u nekom smislu, „najbolje uklapa“ (engl. *„best fit“*) u podatke.

Njemački znanstvenik Karl Gauss predložio je procjenu parametara β_0 i β_1 u jednadžbi (2-2) kako bi se minimalizirao zbroj kvadrata vertikalnih odstupanja. Ovaj kriterij za procjenu koeficijenata regresije naziva se metodom najmanjih kvadrata.

Koristeći jednadžbu (2-2), *n* promatranja u uzorku možemo izraziti kao

$$y_i = \beta_0 + \beta_1 x_i + \in_i \quad , \qquad i = 1, 2, ..., n \tag{2-3}$$

te zbroj kvadrata odstupanja promatranja od pravog (stvarnog) pravca regresije je opisan sljedećim izrazom (funkcijom cilja):

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{2-4}$$

Prema [1], estimatori najmanjih kvadrata do β_0 i β_1 , označenih kao $\hat{\beta}_0$ i $\hat{\beta}_1$, moraju minimizirati funkciju cilja. Kako je funkcija cilja konveksna minimum je moguće odrediti kao nultočke parcijalnih derivacija funkcije cilja po estimiranim parametrima

$$\frac{\partial L}{\partial \beta_0}|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2-5}$$

$$\frac{\partial L}{\partial \beta_1}|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Pojednostavljivanjem prethodne dvije jednadžbe dobivamo:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_1^2 = \sum_{i=1}^{n} y_i x_i \tag{2-6}$$

Jednadžbe (2-6) nazivaju se normalnim jednadžbama najmanjih kvadrata. Rješenje za normalne jednadžbe rezultiraju estimatorima najmanjih kvadrata $\hat{\beta}_0$ i $\hat{\beta}_1$.

Estimatori najmanjeg kvadrata odsječka na osi i nagiba kod modela jednostavne linearne regresije su

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2-7}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \tag{2-8}$$

gdje je $\bar{y} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} y_i$, te $\bar{x} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} x_i$.

Stoga je procijenjeni (estimirani) pravac regresije

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{2-9}$$

Bitno da svaki promatrani par zadovoljava izraz

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \ , \qquad i = 1, 2, \dots, n$$

gdje je $e_i = y_i - \hat{y}_i$ ostatak (rezidua). Ostatak opisuje pogrešku u prilagodbi modela do *i*-tog promatranja y_i.

2.1.1. Koeficijent determinacije (odlučnosti) R^2

Pri provjeri adekvatnosti (optimalnosti) regresijskog modela potrebno je postaviti nekoliko pretpostavki.

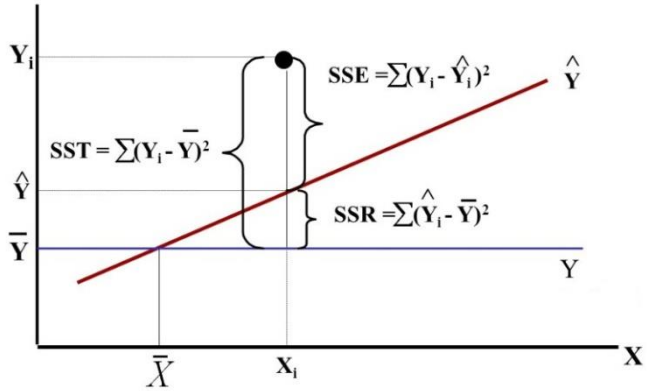
- Procjena parametara modela zahtijeva pretpostavku da su pogreške nekorelirane (nepovezane) slučajne varijable sa srednjom vrijednosti nula i konstantnom varijancom.
- Ispitivanja hipoteze i procjena intervala zahtijevaju da pogreške budu normalno raspoređene.
- Uz to pretpostavlja se da je odabrani model točan, tj; ako koristimo model jednostavne linearne regresije, pretpostavlja se da se taj fenomen stvarno ponaša linearno.

S obzirom na to da bi uvijek trebalo razmotriti valjanost ovih pretpostavki i provesti analize kako bi se ispitala primjerenost modela, jedna od korisnih metoda za to je određivanje koeficijenta determinacije (R^2) - omjer zbroja kvadrata.

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{2-10}$$

gdje je:

- $SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ (engl. *regression sum of squares*), suma kvadrata odstupanja regresijskih vrijednosti od prosjeka
- $SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ (engl. *error sum of squares*), suma kvadrata odstupanja stvarnih vrijednosti zavisne varijable od pripadajućih regresijskih vrijednosti
- $SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$ (engl. *total corrected sum of squares*), suma kvadrata odstupanja stvarnih vrijednosti zavisne varijable od njezina prosjeka

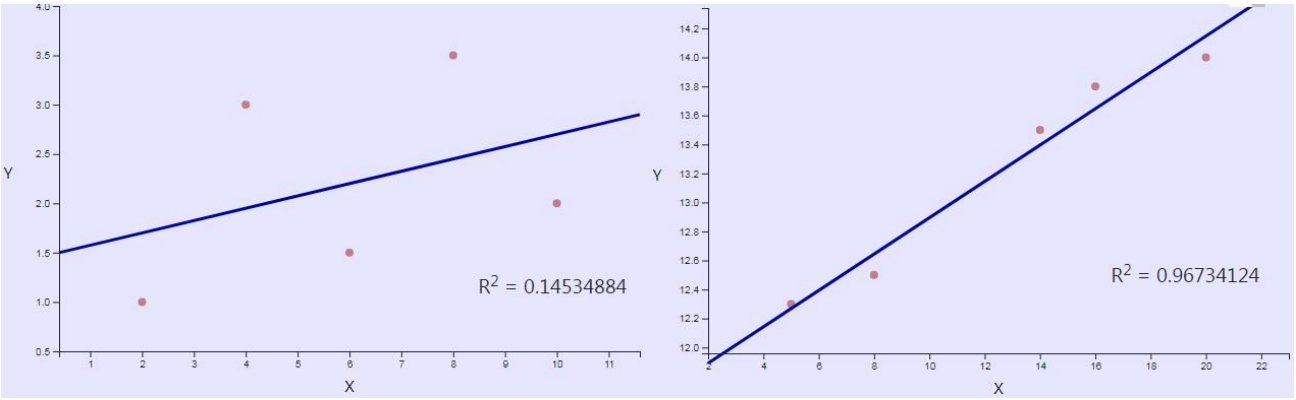


Sl. 2.2. *Grafički prikaz SS_T, SS_E, SS_R.*

Koeficijent determinacije R^2 se često koristi za prosudbu o adekvatnosti regresijskog modela, te poprima vrijednosti iz sljedećega intervala:

$$0 \leq R^2 \leq 1$$

Što je vrijednost R^2 koeficijenta bliža jedinici to je regresijski model reprezentativniji, te ukazuje na to da linearni model objašnjava veliki dio rasipanja u eksperimentalnim vrijednostima zavisne varijable. R^2 koeficijent se izražava u postotcima, stoga rezultate iz primjera sa slike 2.3. možemo izraziti kao R^2=14,53% (lijevo) i R^2=96,73% (desno).



Sl. 2.3. *Raspršenost podataka s nižim R^2 koeficijentom (lijevo) i s R^2 koeficijentom bližim 1 (desno).*

2.2. Višestruka linearna regresija

Mnoge primjene regresijske analize uključuju situacije kada postoji više od jednog regresora ili nezavisnih varijabli. Regresijski model koji sadrži više od jedne nezavisne varijable je nazvan modelom višestruke regresije. Prema [1], općeniti model gdje ovisna varijabla Y može biti povezana s k neovisnih varijabli glasi:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{2-11}$$

te se takav model naziva višestruki linearni regresijski model s k nezavisnih varijabli. Parametri β_j, $j=0,1,\ldots,k$, nazivaju se koeficijenti regresije. Parametar β_j predstavlja očekivanu promjenu zavisne varijable Y po jedinici promjene x_j kada su sve preostale nezavisne varijable x_i ($i \neq j$) konstante. Višestruki linearni regresijski modeli često se koriste pri aproksimaciji funkcija. Odnosno, pravi funkcionalni odnos između Y i x_1, x_2,..., x_k je nepoznat, ali u određenim rasponima neovisnih varijabli, model linearne regresije je adekvatna aproksimacija.

Metoda najmanjih kvadrata može se koristiti za procjenu koeficijenata regresije u višestrukom regresijskom modelu, jednadžba (2-11). Pretpostavimo da je dostupno $n>k$ podataka (mjerenja) i neka x_{ij} označava i-to mjerenje varijable x_j, tj:

$$(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i), \qquad i = 1,2,\ldots,n, \qquad n > k$$

te tako dobivamo:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \\ &= \beta_0 + \textstyle\sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i \qquad\qquad i = 1,2,\ldots,n \end{aligned} \tag{2-12}$$

Uobičajeno je da se podaci za višestruku linearnu regresiju prikazuju u tablici (Tab 2.1).

Tab. 2.1. *Prikaz podataka višestruke regresije.*

y	x_1	x_2	$\cdots$	x_k
y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

Funkcija najmanjih kvadrata, gdje želimo minimizirati L s obzirom na $\beta_0, \beta_1, \dots, \beta_k$, glasi:

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \tag{2-13}$$

Prema [2], estimatori najmanjih kvadrata $\beta_0, \beta_1, \dots, \beta_k$ moraju zadovoljiti sljedeće:

$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) = 0 \tag{2-14}$$

$$\left.\frac{\partial L}{\partial \beta_j}\right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) x_{ij} = 0 \qquad j = 1,2, \dots, k \tag{2-15}$$

Pojednostavljivanjem jednadžbi (2-14) i (2-15) dobivaju je normalne jednadžbe najmanjih kvadrata.

$$\begin{aligned}
n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} \quad &+ \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} \quad + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} \quad = \sum_{i=1}^{n} y_i \\
\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 \quad &+ \hat{\beta}_2 \sum_{i=1}^{n} x_{i1} x_{i2} \quad + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{i1} x_{ik} \quad = \sum_{i=1}^{n} x_{i1} y_i \\
\vdots \qquad\qquad &\vdots \qquad\qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\qquad\quad \vdots \\
\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} x_{ik} x_{i1} &+ \hat{\beta}_2 \sum_{i=1}^{n} x_{ik} x_{i2} \quad + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 \quad = \sum_{i=1}^{n} x_{ik} y_i
\end{aligned} \tag{2-16}$$

Postoji p=k+1 normalnih jednadžbi, po jedna za svaki nepoznati koeficijent regresije.

Rješenja normalnih jednadžbi su estimatori najmanjih kvadrata regresijskih koeficijenata, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$. Normalne jednadžbe mogu se rješavati bilo kojom metodom prikladnom za rješavanje sustava linearnih jednadžbi.

2.2.1. Matrični pristup višestrukoj linearnoj regresiji

Prema [2] mnogo je prikladnije izraziti matematičke operacije s pomoću matričnog zapisa kod postavljanja modela višestruke regresije. Model prikazan jednadžbom (2-12) je sustav *n* jednadžbi koji se može izraziti preko matričnog zapisa kao:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{2-17}$$

gdje je:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Matrica $\boldsymbol{X}$ još se naziva matrica regresora, te njen prvi stupac sadrži jedinice dok su ostali stupci promatrane vrijednosti nezavisnih varijabli $x_1, x_2, \ldots, x_k$.,dok je $\boldsymbol{y}$ (n×1) vektor promatranih podataka, $\boldsymbol{\beta}$ je (p×1) vektor koeficijenata regresije, te je $\boldsymbol{\epsilon}$ (n×1) vektor slučajnih pogrešaka.
Vektor estimatora najmanjih kvadrata $\widehat{\boldsymbol{\beta}}$ dobivamo traženjem $\boldsymbol{\beta}$ iz jednadžbe:

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0 \tag{2-18}$$

gdje je $L = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$.

Rješavanje (2-18) rezultira normalnim jednadžba metode najmanjih kvadrata u matričnom obliku:

$$\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y} \tag{2-19}$$

Množenjem obje strane jednakosti kod (2-19) s inverzom od $\boldsymbol{X}'\boldsymbol{X}$ dobivamo rješenja normalnih jednadžbi, estimatore najmanjih kvadrata.

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{2-20}$$

Također možemo zapisati procijenjeni (estimirani) model u matričnom obliku kao:

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{2-21}$$

U praksi se izračuni višestruke regresije gotovo uvijek izvode s pomoću računala.

2.2.2. Polinomni regresijski modeli

U statistici, prema [1], polinomna regresija oblik je regresijske analize u kojoj se odnos između nezavisne varijable *x* i zavisne varijable *Y* modelira kao polinom *n*-tog stupnja u ovisnosti o *x*. Polinomska regresija odgovara nelinearnom odnosu između vrijednosti *x* i odgovarajuće očekivane aritmetičke sredine od *Y*, označeno s $E(Y \mid x)$.

Iako polinomna regresija odgovara nelinearnom modelu, ona je kao problem statističke estimacije linearna, u smislu da je regresijska funkcija $E(Y|x)$ linearna u nepoznatim parametrima koji se procjenjuju iz podataka. Zbog toga se polinomna regresija smatra posebnim slučajem višestruke linearne regresije. Općenito, svaki regresijski model koji ima linearne parametre β linearan je regresijski model, bez obzira na funkciju kojom aproksimira.

Linearni model $\boldsymbol{Y = X\beta + \epsilon}$ opći je model koji se može koristiti za uklapanje u bilo koji odnos koji je linearan u nepoznatim parametrima β, uključujući polinomne regresijske modele. Kao primjer možemo zapisati polinom drugog stupnja preko jedne varijable kao:

$$Y = \beta_0 + \beta_1 x_1 + \beta_{11} x^2 + \epsilon \tag{2-22}$$

te polinom drugog stupnja preko dvije varijable:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon \tag{2-23}$$

Polinomni regresijski modeli često se koriste kada je odgovor grafa funkcije krivulja jer se mogu primijeniti opća načela višestruke regresije.

2.3. Područja primjene

Općenito regresijska analiza ima široku primjenu u raznim granama znanosti - od medicine, biologije, sporta, marketinških i ekonomskih istraživanja, fizike pa do, sve popularnijeg, strojnog učenja (engl. *machine learning*).

Primjeri uporabe iz prakse: [3]

- unutar ekonomije koristi se za razumijevanja veze između troškova oglašavanja i prihoda,
- u sportu se mjeri učinkovitost različitih režima treninga na igračevu izvedbu,
- istraživanja u medicini kako bi se razumio odnos doziranja lijeka i krvnog tlaka pacijenta,
- poljoprivredni znanstvenici koriste za mjerenje učinkovitosti gnojiva i vode na prinos usjeva

3. ALATI ZA IZRADU WEB-APLIKACIJE

U okviru završnog rada bilo je potrebno napraviti web-aplikaciju koja će rješavati zadatke s pomoću metode najmanjih kvadrata.

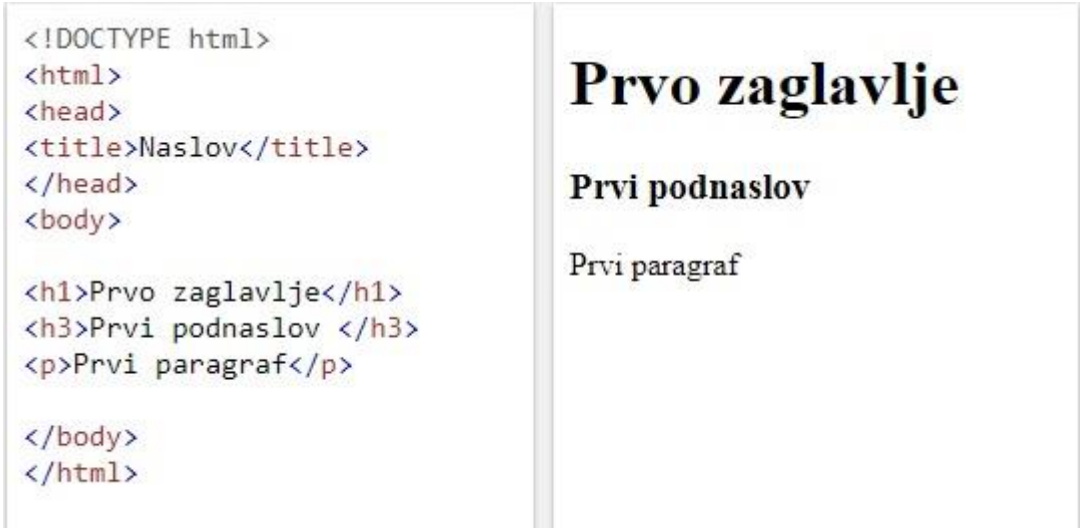
3.1. VisualStudio Code

VisualStudio Code je prilagodljiv, jednostavan za korištenje i besplatan editor za pisanje kôda tj. IDE (engl. *Integrated Development Environment*) napravljen od strane Microsofta.

3.2. HTML I CSS

HTML (engl. *Hypertext Markup Language*) je standardni i osnovni način označavanja teksta koji će biti prikazan u web-pregledniku, opisujući strukturu web-stranice ili web-aplikacije.[4]

Elementi HTML-a su prikazani s pomoću oznaka (engl. *tags*) te s pomoću njih HTML elementi „govore“ web-pregledniku kako prikazati sadržaj. HTML oznakama stranice na internetu mogu prikazivati tekst i multimedijske sadržaje u različitim veličinama, bojama i oblicima.



Sl. 3.1. *Primjer oblikovanja web-stranice u HTML-u te izgled u pregledniku.*

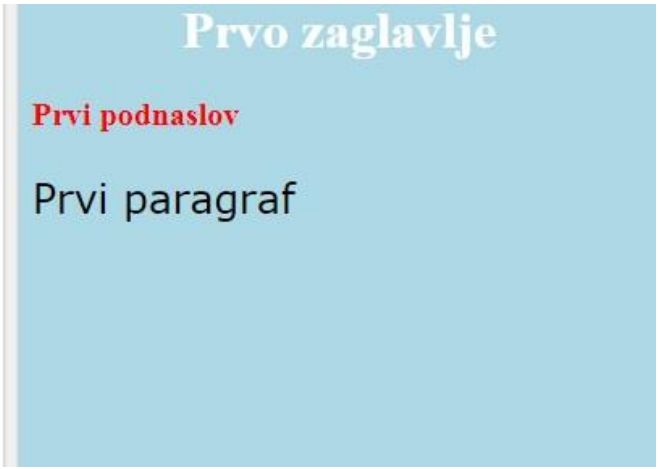
Kao što se vidi na slici 3.1. web-preglednici ne prikazuju HTML oznake već ih koriste za interpretaciju sadržaja unutar preglednika. Sve što želimo prikazati u pregledniku pišemo unutar *<body>..</body>* oznaka. Još neke bitne oznake s opisom prikazane su na slici 3.2.

HTML oznake i parametri	Opis oznaka i parametra
<html>	Označava početak web stranice
<head>	Označava zaglavlje
<title>	Ime dokumenta koje se pojavljuje u web pregledniku
<meta>	Ova oznaka nema završne oznake, a upotrebljava se za definiranje obilježja web stranice
<body>	Označava tijelo (sadržaj) dokumenta
<p>	Označava odlomak
 	Ubacuje prijelom teksta u novi red
<b>	Podebljana slova
<i>	Ukošena slova
<u>	Podvučena slova
<p align="left">	Poravnanje odlomka lijevo

Sl. 3.2. *Najčešće oznake unutar HTML-a.*

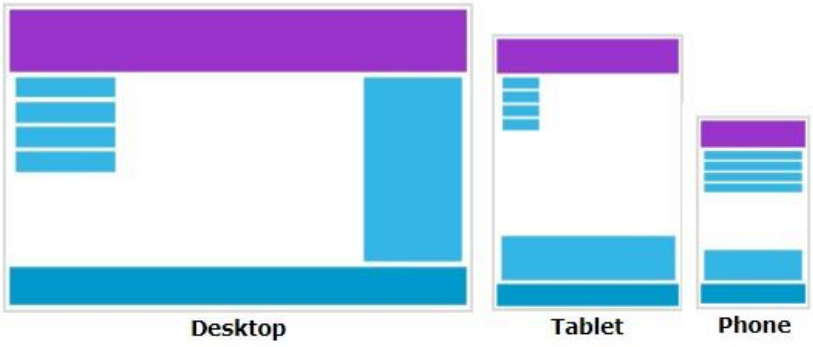
Uz HTML se najčešće koristi i CSS (engl. *Cascading Style Sheets*) za lakše oblikovanje sadržaja prikazanog unutar web-preglednika. CSS opisuje kako će dokument pisan HTML-om biti prezentiran te kako će njegove oznake biti prikazane.

```
body {
  background-color: lightblue; }

h1 {
  color: white;
  text-align: center; }

h3 {
    font-size: 20px;
    color: red; }

p {
  font-family: verdana;
  font-size: 25px; }
```



Sl. 3.3. *Kôd CSS-a te izgled web-stranice nakon primijene CSS-a.*

Na slici 3.3. prikazan je primjer oblikovanja web-stranice s pomoću CSS-a. CSS je napravljen tako da omogućuje odvajanje oblikovanja od sadržaja, uključujući određivanje izgleda web-stranice i pozicioniranja elemenata unutar nje, te boja i fontova za prikaz.[5]
CSS je također zaslužan za prilagođeno prikazivanje sadržaja web-stranice za sve širine ekrana (laptopi, tableti, mobiteli) s pomoću responzivnog dizajna.



Sl. 3.4. *Responzivni dizajn na različitim uređajima.*

3.3. JavaScript

Uz HTML i CSS, JavaScript spada među osnovne tehnologije korištene za izradu sadržaja na internetu. JavaScript omogućava interaktivnost web-stranica te je glavni dio web-aplikacija.

JavaScript je bio zamišljen kao programski jezik koji će se izvršavati na klijentskoj strani (engl. *client-side*), tj. unutar web-preglednika, ali se proširio i koristi se i na serverskoj strani (engl. *server-side*) u okruženjima bez web-preglednika kao što je Node.js, serverska JavaScript platforma.
Budući da je JavaScript tako dobro prihvaćen kao programski jezik pojavila se potreba za formiranje standarda, odnosno specifikacija pod nazivom ECMAScript. Sada je već aktualno 11. izdanje ECMAScript-a. [6]

Varijable u JavaScriptu služe kao spremnici za pohranu vrijednost. Do 2015. godine ključna riječ *var* je bila jedini način deklariranja varijabli u JavaScriptu. Iste godine verzijom ECMAScripta6 dopušta se uporaba još nekih ključnih riječi za deklaraciju varijabli:

- *const* za definiranje varijable čija vrijednost ne može biti promijenjena (konstante),
- *let* za definiranje varijabli s ograničenim djelokrugom (engl. *scope*).

```
var x = 10;
// Tu je  x = 10
{
  var x = 2;
  // Tu je   x = 2
}
// Tu je   x = 2
```

```
var x = 10;
// Tu je   x = 10
{
  let x = 2;
  // Tu je   x = 2
}
// Tu je   x = 10
```

Sl. 3.5. *Razlike pri korištenju varijabli 'var' i 'let'.*

Bitan dio JavaScripta su i uvjetna grananja, te s pomoću njih se upravlja programom i utvrđuje se hoće li se određeni dio kôda izvršiti ili neće. Postoji više vrsta uvjetnih grananja, kao što su:

- *if* – izvršava se ako je uvjet istinit (engl. *true*)
- *if.. else* – ako je uvjet *if* neistinit (engl. *false*) izvršava se dio kôda specificiran pod *else*
- *if...else if...else* – ako je prvi uvjet *if* neistinit specificira se novi uvjet pod *else if*, te ako je on istinit izvršava se taj dio kôda, a ako je neistinit izvršava se dio kôda pod *else*

```
if (uvjet1) {
   // naredbe koje će se izvršiti ako je vrijednost uvjeta1 true;
} else if (uvjet2) {
   // naredbe koje će se izvršiti  je vrijednost uvjeta2 true;
} else if (uvjet3) {
   // naredbe koje će se izvršiti  je vrijednost uvjeta3 true;
} else {
  // naredbe se izvršavaju ako su i uvjet1, uvjet2 i uvjet3 false;
}
```

Sl. 3.6. *if ... else if ... else uvjetno grananje.*

JavaScript je poslužio i kao podloga za pisanje raznih biblioteka (engl. *library*) čiji je cilj olakšavanje izrade JavaScript aplikacija. Biblioteke su skup funkcija pisanih JavaScript programskim jezikom.[6]
Neke od najpoznatijih JavaScript biblioteka su: D3.js, jQuery, React.

3.3.1. D3.js

Jedna od biblioteka JavaScripta je i D3.js (engl. *Data-Driven Documents*) biblioteka koja služi za izradu dinamičke, interaktivne vizualizacije podataka unutar web-preglednika.
Pri izradi vizualizacije podataka koristi vektorsku grafiku tj. SVG (engl. *Scalable Vector Graphic*), HTML i CSS. [7]

Ugradnjom unutar web-aplikacije D3.js datoteka koristi već ugrađene JavaScript funkcionalnosti za odabir elemenata, stvaranje SVG objekata i uređivanje istih, može se također dodati i premještanje (tranzicija) te neki dinamički efekti. Iako sve što može datoteka D3.js može se postići i CSS-om, velikom skupinom podataka lakše je manipulirati uporabom datoteke.

Podatci koji se unose u D3.js mogu biti raznih formata, najčešći su JSON ili CSV (engl. *coma-separated values*). Pri korištenju izrađene web-aplikacije za računanje metodom najmanjih kvadrata podatci se unose u CSV formatu.

D3.js upravlja dinamičkim podatcima prihvaćanjem općeg obrasca ažuriranja (engl. *general update pattern*). Unutar toga obrasca dodavanje i izmjena elemenata obavlja se metodama: [8]

- *enter()* – sadrži „spremnik“ za bilo koji novi element
- *update()* – sadrži postojeće elemente kojima su ažurirani podatci
- *exit()* – sadrži preostale elemente koji nisu ažurirani, tj. koji su višak.

```
let setChart = (selection, {dataset}) => {
 let dots = selection.selectAll(".dot")
  .data(dataset);
  dots.join(
   enter => enter
       .append("circle")
       .attr("cx", (d) => x(d.x))
       .attr("cy", (d) => y(d.y))
       .attr("r", 4)
       .attr("fill", "#69b3a2")
       .attr("opacity", 0.7)
       .attr("id", "dots")
       .attr("clip-path", "url(#clip)"),
    update => update
       .append("circle")
       .attr("cx", (d) => x(d.x))
       .attr("cy", (d) => y(d.y)),
    exit => exit.exit().remove()
    );
```

Sl. 3.7. *Kôd iz aplikacije gdje je primijenjen opći obrazac ažuriranja.*

3.4. FileAPI

FileAPI pruža sučelje za programiranje aplikacija (engl. *Application Programming Interface*) za prikazivanje objekata unutar datoteke u web aplikacijama, tj. to je standard koji upravlja datotečnom pohranom na klijentskoj strani i podržavaju ga svi popularni web preglednici.
FileAPI opisuje kako će interakcije s datotekama biti provedene, za čitanje informacija te za čitanje datoteka također s ciljem učitavanja. [9]

U okviru izrade web aplikacije FileAPI je korišten za privremenu pohranu datoteka. Radi na principu spremanja datoteke u priručnu (engl. *cache*) memoriju.
Objekt *FileReader* dopušta web-aplikacijama asinkrono učitavanje datoteka spremljenih na korisnikovom računalu.

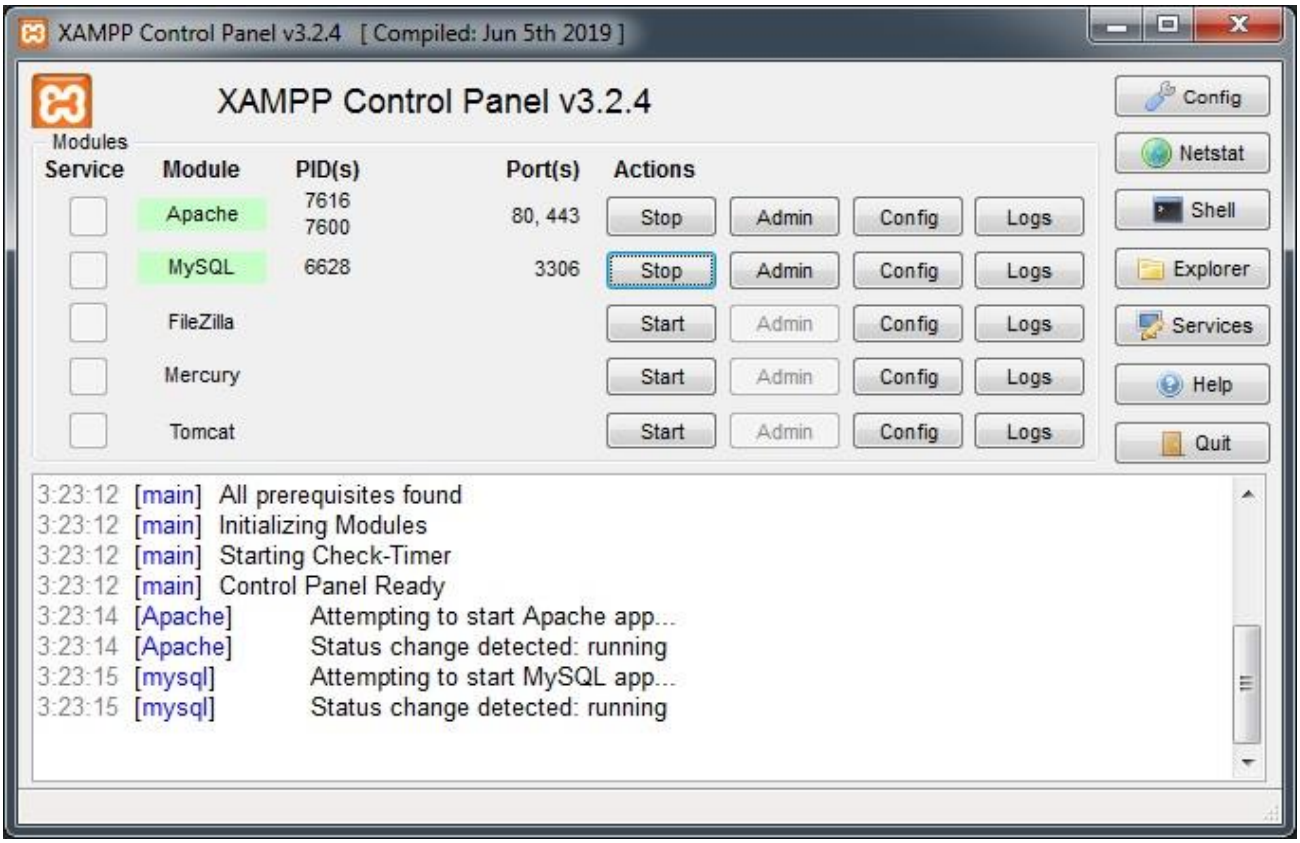
```
function getAsUrl(event) {
  let files = event.target.files[0];
  let reader = new FileReader();

  reader.onerror = errorHandler;

  reader.onload = (function (theFile){
    return function(event) {
      drawChart(event.target.result);
    };
  }) (files);

  reader.readAsDataURL(files);
  processData(files);
}
```

Sl. 3.8. *Glavna funkcija za učitavanje datoteke pomoću FileAPI-a.*

Kôd na slici 3.8. prikazuje JavaScript funkciju unutar koje definiramo varijable potrebne za inicijalizaciju FileAPI-ja.

3.5. XAMPP

XAMPP (engl. *Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P)*) je besplatna i multi-platforma otvorenoga kôda (engl. *open-source*) koja služi kao lokalni web-server. Pomoću XAMPP-a može se otvoriti web-stranica ili web-aplikacija lokalno na računalu te se koristi pri izradi web-sadržaja za testiranje funkcionalnosti.

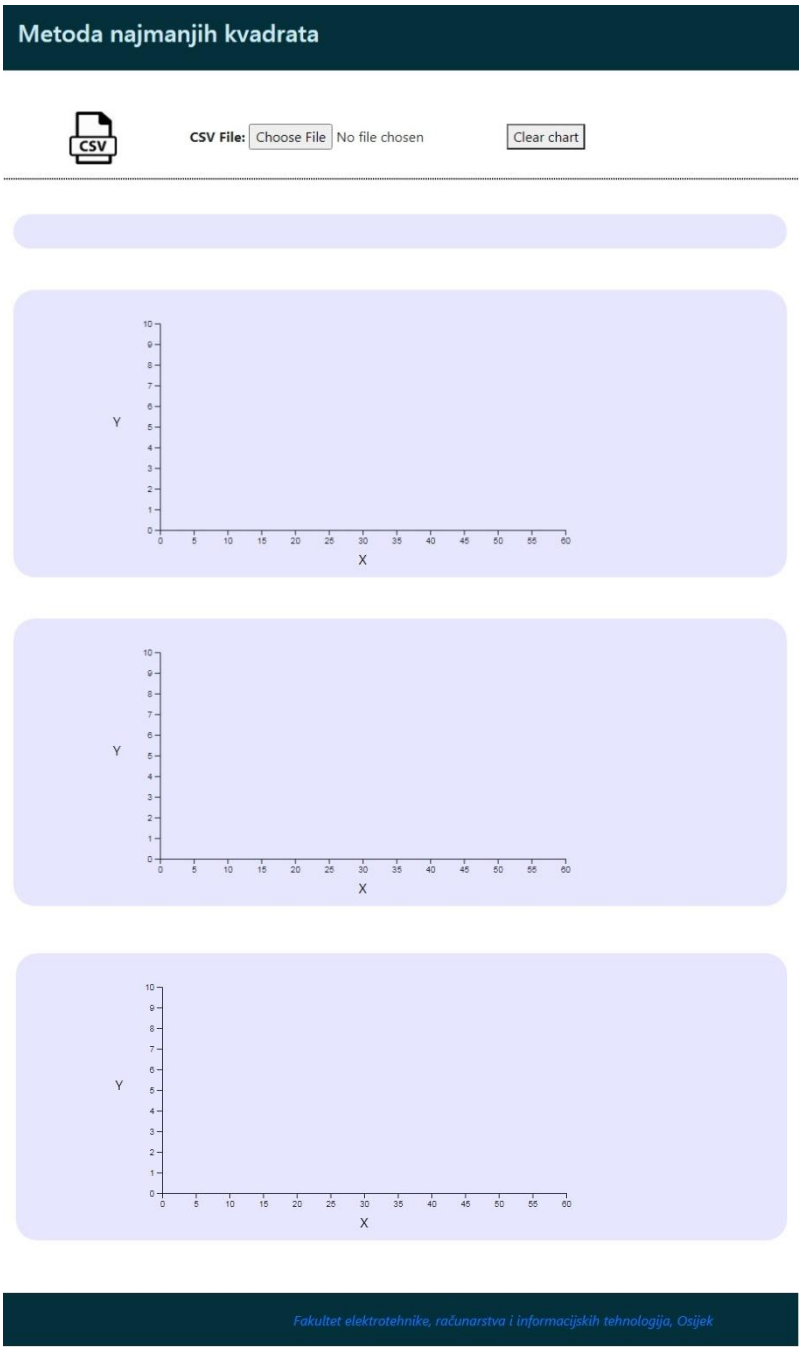


Sl. 3.9. *Sučelje XAMPP-a.*

Kako bi web-sadržaj bio uspješno prikazan u pregledniku nužno je da su prva dva servisa (*Apache* i *MySQL*) na sučelju XAMPP-a pokrenuta (slika 3.9.), te je potrebno HTML, CSS, JavaScript i eventualne druge datoteke potrebne za rad smjestiti na odgovarajuće mjesto. Kod XAMPP-a putanja za to je sljedeća: *disk* na kojemu je program instaliran → *XAMPP* → *htdocs*.

4. PROGRAMSKA IMPLEMENTACIJA

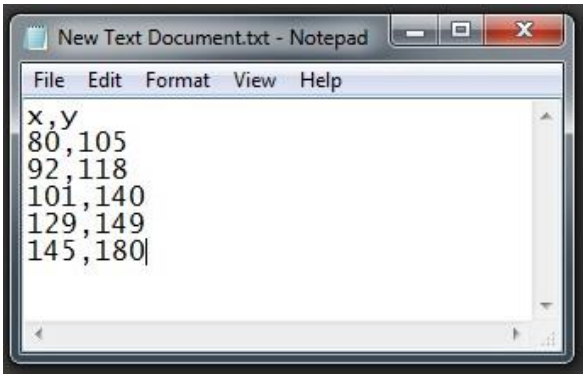
U ovom poglavlju bit će opisana uporaba izrađene web-aplikacije, a za koju je dana teorijska osnova u prethodnim poglavljima.



Sl. 4.1. *Početni izgled web-aplikacije.*

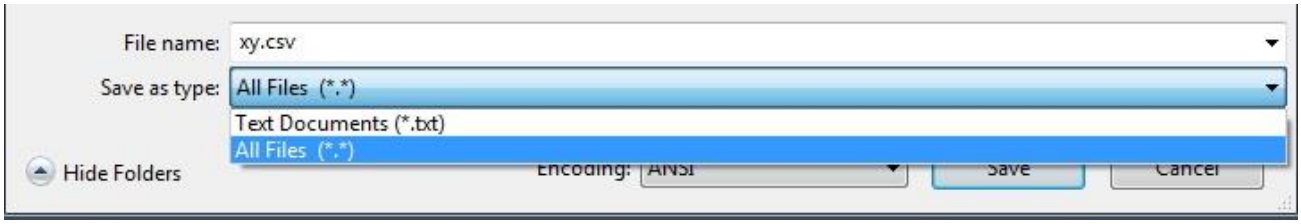
Na slici 4.1. je prikazan početni izgled, te unutar aplikacije klikom na prvu ikonu skinuti će se na računalo .csv datoteka koja može poslužiti kao predložak za unos podataka u aplikaciju.
Pokraj se nalaze dva gumba; prvi za odabir datoteke koja sadrži podatke koji se žele prikazati, te drugi gumb koji služi za brisanje prikazanih podataka prije učitavanja novih.
Te nakon toga ispod horizontalne linije nalazi se mjesto predviđeno za ispis koja od linearnih regresija najbolje opisuje učitane podatke s obzirom na R^2 koeficijent. Ispod toga se nalaze 3 grafa, gdje je prvi predviđen za prikaz linearne regresije, drugi graf za prikaz kvadratne regresije te treći graf za prikaz kubne regresije.

CSV datoteka se najlakše kreira iz običnog tekst editora (*Notepad*) dodajući datoteci ekstenziju .csv. Može se također kreirati s pomoću Microsoft Office Excela.



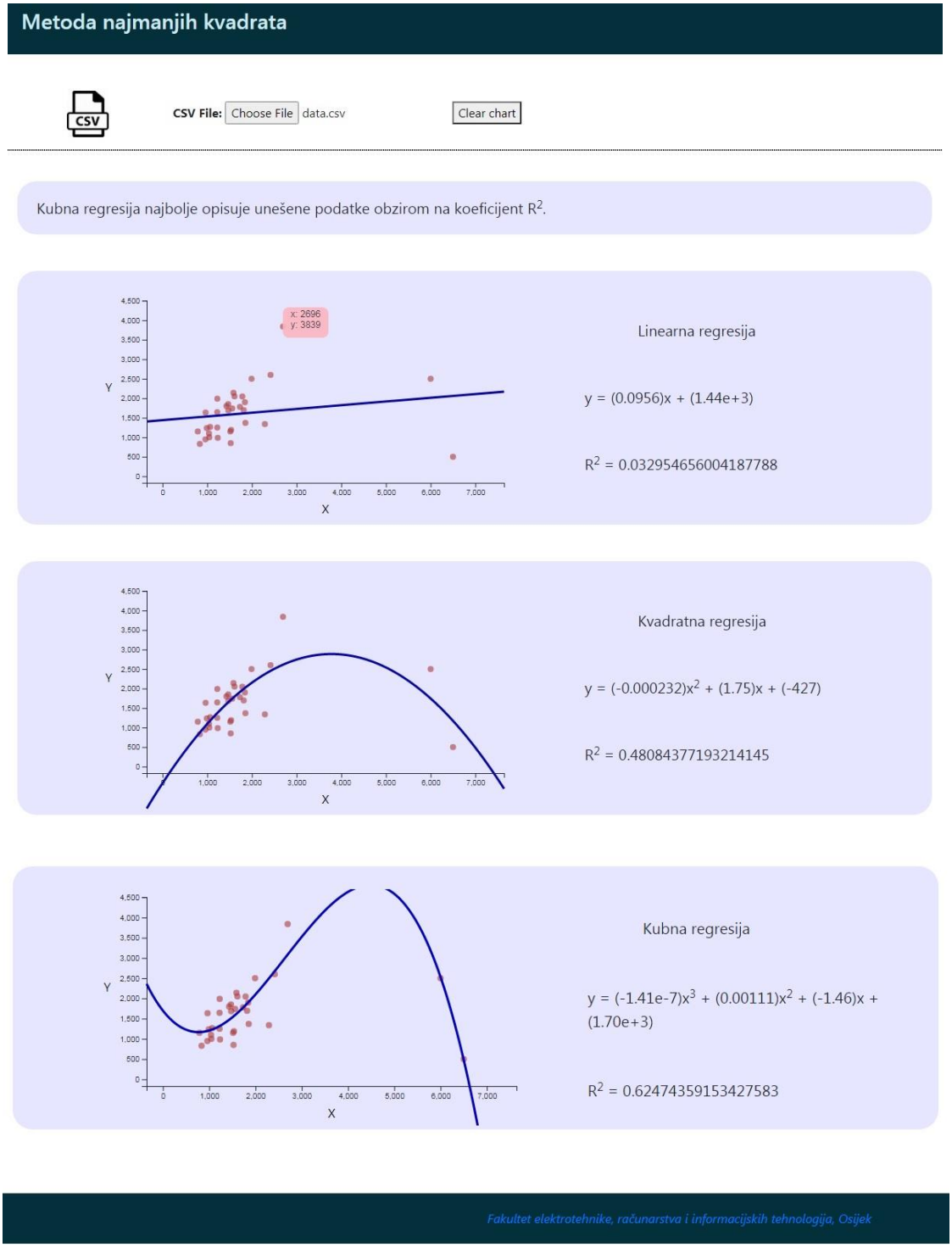
Sl. 4.2. *Format podataka za unos u aplikaciju.*

Kao što se vidi na slici 4.2. prvi red je predodređen za zaglavlje, te se podatci za svaku koordinatu unose u poseban stupac. Vrijednosti moraju biti odvojene zarezom.
Datoteku iz *Notepada* spremamo: *File* → *Save As* → *Save as type :* mijenjamo na *All Files* ,
File name → nazovemo datoteku kako želimo uz obvezan dodatak .csv ekstenzije.



Sl. 4.3. *Spremanje CSV datoteke.*

Nakon učitavanja podataka dobiju se tri iscrtana grafa i pokraj svakoga njegov opis.



Sl. 4.4. *Izgled nakon unosa podataka.*

Prema slici 4.4. na grafovima, nakon unosa podataka, pojavljuju se točke u koordinatnom sustavu koje prikazuju koordinate točke te se prikazuje regresijski pravac. Pokraj grafa se ispisuju naziv regresije za svaki graf, jednadžba pravca za y te R^2 koeficijent.
Jednadžba krivulje ispisana je u formatu:

$$y = ax + b \tag{4-1}$$

gdje je a estimirani koeficijent smjera pravca, a b estimirani odsječak na y-osi, kao i što je navedeno u drugomu poglavlju.

Izrađena web-aplikacija može prikazati podatke iz sva četiri kvadranta koordinatnog sustava, te nema nikakva ograničenja pri rasponu unesenih podataka, tj. brojevi na koordinatnim osima se prilagođavaju minimalnim i maksimalnim unesenim vrijednostima. Također pri prelasku strelice miša preko svake točke prikazuje se oblačić s ispisanim x i y koordinatama (slika 4.4.).

5. ZAKLJUČAK

U ovome radu cilj je bio izraditi web-aplikaciju za aproksimaciju podataka metodom najmanjih kvadrata. Za izradu su korišteni HTML, CSS, JavaScript te njegova biblioteka D3.js.
Na početku rada opisan je teorijski dio matematike koji je korišten kasnije pri izradi web-aplikacije. Pogodnost ove aplikacije je što se mogu procesirati velike količine podataka te aplikacija izračunava i ispisuje R^2 koeficijent pa s pomoću njega uspoređuje koja regresija najbolje opisuje unesene podatke (linearna, kvadratna ili kubna).

Postoji mogućnost daljnjeg proširenja ove aplikacije na neki od nelinearnih modela, poput eksponencijalnog ili logaritamskog.

Cijeli kôd aplikacije bit će priložen na CD-u.

LITERATURA

[1] D. C. Montgomery, G.C. Runger : „*Applied statistics and probability for engineers*“, 6. izdanje, John Wiley & Sons, SAD, 2013.

[2] R. Scitovski : *„Numerička matematika“* , Odjel za matematiku Sveučilišta u Osijeku, 2004.

[3] Primjene linearne regresije u stvarnom životu [rujan 2020.]
https://www.statology.org/linear-regression-real-life-examples/

[4] HTML i njegove oznake [kolovoz 2020.]
https://www.w3schools.com/html/

[5] CSS [kolovoz 2020]
https://www.w3schools.com/css/default.asp

[6] JavaScript [kolovoz 2020.]
https://www.w3schools.com/js/

[7] D3.js billioteka [rujan 2020.]
https://github.com/d3/d3/blob/master/API.md

[8] Opći obrazac ažuriranja biblioteke D3.js [rujan 2020.]
https://www.freecodecamp.org/news/how-to-work-with-d3-jss-general-update-pattern-8adce8d55418/

[9] FileReader API [rujan 2020.]
https://w3c.github.io/FileAPI/#blobreader-task-source

SAŽETAK

Naslov: Web-aplikacija za aproksimaciju podataka metodom najmanjih kvadrata

U ovom radu predstavljena je web-aplikacija za izračun te ispis grafa linearne, kvadratne i kubne regresije primjenom metode najmanjih kvadrata. Tehnologije korištene pri izradi su HTML, CSS, JavaScript i biblioteka od JavaScripta D3.js

Ključne riječi: D3.js, metoda najmanjih kvadrata, web-aplikacija

ABSTRACT

Title: Web application for data approximation using least square method

This paper presents web application for calculating and plotting chart with linear, quadratic and cubic regression using least square method. Programming languages which were used are HTML, CSS, JavaScript with its D3.js library.

Keywords: D3.js, least square method, web application

PRILOZI

Prilog 1.DVD

- Web-aplikacija za aproksimaciju podataka metodom najmanjih kvadrata.pdf
- Web-aplikacija za aproksimaciju podataka metodom najmanjih kvadrata.docx
- Programsko rješenje web-aplikacije